

Deviations in the Zipf and Heaps laws in natural languages

Bochkarev V., Lerner E., Shevlyakova A.

Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

This paper is devoted to verifying of the empirical Zipf and Hips laws in natural languages using Google Books Ngram corpus data. The connection between the Zipf and Heaps law which predicts the power dependence of the vocabulary size on the text size is discussed. In fact, the Heaps exponent in this dependence varies with the increasing of the text corpus. To explain it, the obtained results are compared with the probability model of text generation. Quasi-periodic variations with characteristic time periods of 60-100 years were also found. © Published under licence by IOP Publishing Ltd.

<http://dx.doi.org/10.1088/1742-6596/490/1/012009>
